Ewa FIGIELSKA[1]

EWOLUCYJNE METODY UCZENIA UKRYTYCH MODELI MARKOWA

Streszczenie

Ukryte modele Markowa stanowią narzędzie modelowania statystycznego wykorzystywane do analizy i przewidywania zjawisk o charakterze sekwencji zdarzeń występujących na przykład w rozpoznawaniu mowy i gestów oraz modelowaniu sekwencji biologicznych. Aby ukryty model Markowa mógł z powodzeniem zostać zastosowany w praktyce, konieczne jest określenie jego topologii i wyznaczenie wartości jego parametrów. Istniejące metody klasyczne nie zawsze są zdolne do dostarczenia wystarczająco dobrych modeli. Dlatego też, w ostatnich latach obserwuje się wzrost zainteresowania możliwością stosowania innych technik, zwłaszcza opartych na mechanizmach stochastycznych. W artykule przedstawione są sposoby wykorzystania w procesie budowy ukrytych modeli Markowa metod ewolucyjnych. Przeprowadzona jest również ocena jakości otrzymywanych w ten sposób modeli.

Abstract

Hidden Markov models (HMMs) are a statistical tool for analyzing and modeling time-series data. They have been successfully used in many areas requiring time-series analysis for example in speech recognition, DNA sequence analysis or forecasts of stock prices. To use a HMM in practice, the topology and the values of its parameters have to be determined. The existing classical methods for HMM training are not always able to provide sufficiently good models. Therefore, in recent years, we observe an increasing interest in developing other methods for HMM training, especially ones involving evolutionary mechanisms. This paper presents how evolutionary methods can be used to build HMMs. The quality of the obtained in this way HMMs is also discussed.

1. WSTĘP

Ukryte modele Markowa (ang. *hidden Markov models*, *HMM*), wprowadzone w drugiej połowie lat sześćdziesiątych ubiegłego wieku przez Bauma i Petriego (Baum i Petrie, 1966), stanowią klasę procesów stochastycznych zdolną do modelowania danych o charakterze sekwencji zdarzeń. Przy ich stosowaniu modelowany

[1] Dr inż. Ewa Figielska jest wykładowcą Warszawskiej Wyższej Szkoły Informatyki.

system jest przedstawiany jako proces Markowa o niewidocznych dla obserwatora stanach, ale z widocznym wyjściem (obserwacją), które jest losową funkcją stanu.

Ukryte modele Markowa ze względu na bogatą strukturę matematyczną, mogą stanowić bazę teoretyczną w szerokim zakresie zastosowań, a użyte we właściwy sposób, dają bardzo dobre rezultaty w praktyce. Stosowane były one w zagadnieniach rozpoznawania mowy (Baker 1975; Rabiner, 1989), pisma ręcznego (El-Yacoubi i in., 1999), obiektów (Cai i Liu, 2001) i gestów (Kim i Chien, 2001), w bioinformatyce (Karplus i in., 1997; Bladi i Brunak, 1998; Cheung, 2004), w biomedycynie (Coast i in., 1990), a także w przewidywaniu cen na rynku papierów wartościowych (Hassan i in., 2007).

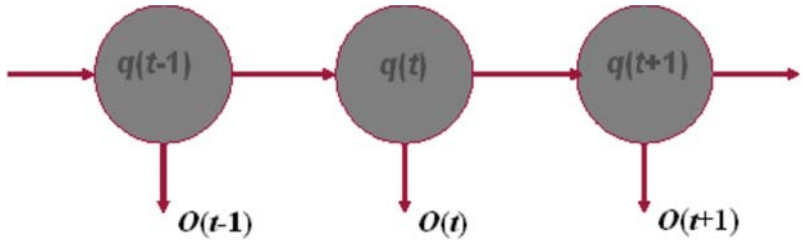
Zasadniczym problemem występującym przy budowie ukrytych modeli Markowa jest taki dobór topologii (liczby stanów i rodzaju powiązań między stanami) oraz wartości parametrów modelu, który zapewnia jego dobre działanie, np. wysoką zdolność rozpoznawania. Klasyczne metody doboru parametrów modelu (algorytmy Bauma-Welcha (Baum i Egon, 1967; Baum et al., 1970, Baum, 1972), metody gradientowe (Jelinek 1976)) nie gwarantują znalezienia ich optymalnych wartości, a także wymagają pewnych wstępnych ustaleń, np. dotyczących topologii modelu. Jakość tworzonych przez nie modeli zależy od przyjętego początkowego oszacowania parametrów modelu, przy czym osiągnięcie oszacowania dobrej jakości wymaga zastosowania procedur o dużej złożoności obliczeniowej (Rabiner i Juang, 1983; Kwong i in., 2001). Dlatego też, w ostatnich latach obserwujemy zainteresowanie poszukiwaniem nowych metod, które pozwalają ominąć ograniczenia metod klasycznych i umożliwiają równoczesne wyznaczanie zarówno optymalnych wartości parametrów jak i topologii modelu.

W następnych rozdziałach niniejszego artykułu przedstawiona zostanie charakterystyka ukrytych modeli Markowa, zasady działania metod ewolucyjnych i sposoby ich zastosowania do oszacowania parametrów ukrytych modeli Markowa, a także ocena jakości uzyskiwanych tymi metodami modeli oraz perspektywy na przyszłość.

2. CHARAKTERYSTYKA UKRYTYCH MODELI MARKOWA

Ukryty model Markowa jest procesem stochastycznym określonym przez dwa powiązane ze sobą mechanizmy: właściwy łańcuch Markowa o pewnej liczbie stanów oraz, związane ze stanami, losowe funkcje. W każdej dyskretnej chwili czasu, proces znajduje się w jednym stanie oraz generowana jest obserwacja przez pewną losową funkcję. W następnej chwili czasu, łańcuch Markowa przechodzi do następnego stanu zgodnie z pewnym określonym dla danego stanu prawdopodobieństwem.

Obserwator widzi tylko wynik działania losowych funkcji, przy czym nie może bezpośrednio obserwować stanów łańcucha Markowa – stąd nazwa „ukryty".



Rys. 1. Architektura ukrytego modelu Markowa, q (t) – stan w chwili t, O (t) – obserwacja w chwili t

2.1. Parametry modelu

Ukryty model Markowa jest opisany za pomocą następujących parametrów (Rabiner i Juang, 1993):

1. Liczba stanów N,
2. Zbiór stanów S, $S = \{S_1, S_2, ..., S_N\}$,
3. Wektor prawdopodobieństw stanu początkowego $\pi = \{\pi_i\}$,
$$\pi = P[q_1 = S_i],\ 1 \le i \le N,$$
gdzie q_1 jest stanem w chwili t=1.
4. Macierz prawdopodobieństw przejść $A = \{a_{ij}\}$,
$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i],$$
gdzie q_t oznacza stan w chwili t, i a_{ij} ma następujące własności:
$$a_{ij} \ge 0,\ 1 \le i, j \le N,$$
$$\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N\ .$$
5. Wektor prawdopodobieństw obserwacji $B = \{b_j = (O)\}$, gdzie $b_j(O)$ jest funkcją losową o postaci (dla modeli o ciągłych gęstościach obserwacji)
$$b_j(O) = \sum_{m=1}^{M} c_{jm} \Theta(O, \mu_{jm}, U_{jm}), \quad 1 \le j \le N\ ,$$
gdzie O jest wektorem obserwacji (modelowany wektor), M jest liczbą składników w funkcji losowej, c_{jm} oznacza współczynniki mieszania spełniające następujące ograniczenia:
$$c_{jm} \ge 0,\ 1 \le j \le N,\ 1 \le m \le M,$$

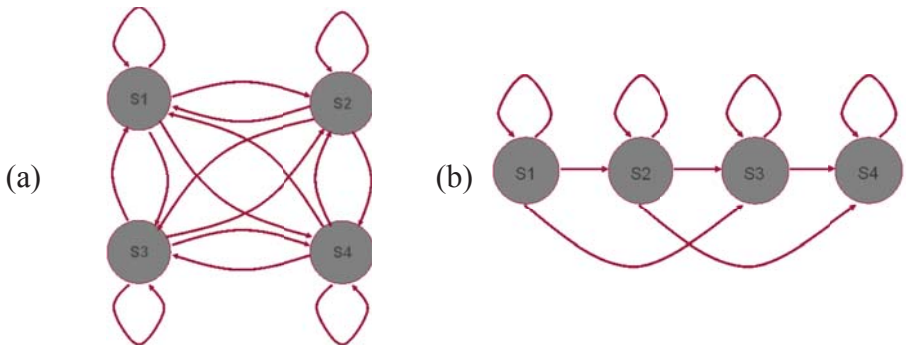
$$\sum_{m=1}^{M} c_{jm} = 1, \quad 1 \le j \le N ,$$

$\Theta(\cdot)$ oznacza dowolny logarytmicznie wklęsły lub eliptycznie symetryczny rozkład gęstości prawdopodobieństwa, np. rozkład normalny z wektorem średnich μ_{jm} i macierzą kowariancji U_{jm}.

2.2. Topologia modelu

Topologia modelu jest określona przez liczbę stanów i rodzaj powiązań między stanami. Można wyróżnić dwa podstawowe typy ukrytych modeli Markowa:

- model ergodyczny (rys. 2a), w którym każdy stan może być osiągnięty z każdego innego stanu,
- model Bakisa (model „od lewej do prawej") (rys. 2b), w którym sekwencja stanów ma tę własność, że wraz ze wzrostem czasu wzrasta numer stanu, tzn. stany przebiegają od strony lewej do prawej.



Rys. 2. Dwa typy ukrytych modeli Markowa: (a) model ergodyczny o 4 stanach, (b) model Bakisa o 4 stanach (linie ciągłe oznaczają możliwe przejścia między stanami)

Model typu „od lewej do prawej" jest odpowiedni dla modelowania sygnałów, których własności zmieniają się w czasie, np. mowy. Wszystkie modele typu „od lewej do prawej" mają tę własność, że nie są w nich dozwolone przejścia do stanów, których numery są mniejsze od numeru stanu aktualnego, $a_{ij} = 0$ dla $j < i$, oraz sekwencja stanów zaczyna się w stanie 1, $\pi_1 = 1$, $\pi_i = 0$, jeżeli $i \neq 1$. Dla ostatniego stanu w modelu „od lewej strony do prawej" współczynniki przejść między stanami spełniają zależności: $a_{NN} = 1$ i $a_{Ni} = 0$ dla $i < N$.

2.3. Specyfikacja i ocena modelu

Aby móc wykorzystać ukryty model Markowa, konieczne jest określenie jego topologii, liczby składników *M* losowej funkcji, wartości elementów macierzy prawdopodobieństw przejść *A*, parametrów funkcji losowych w wektorze obserwacji *B* oraz wektora prawdopodobieństw stanu początkowego. Dla wygody pełny zbiór parametrów ukrytego modelu Markowa oznaczamy przez $\lambda = (A, B, \pi)$. Określenie zbioru wartości parametrów modelu odbywa się przez jego uczenie (trenowanie) z wykorzystaniem pewnego zbioru obserwacji, który traktowany jest jako zbiór danych trenujących.

Gdy dana jest sekwencja obserwacji $O=O_1O_2...O_T$, ocena modelu $\lambda = (A, B, \pi)$ o znanych parametrach następuje przez obliczenie prawdopodobieństwa wygenerowania przez model sekwencji obserwacji *O*. Prawdopodobieństwo to, oznaczane przez $P(O \mid \lambda)$, obliczane jest za pomocą następującej procedury (*forward procedure*, Rabiner i Juang (1993)):

Niech $\alpha_t(i) = P(O_1O_2 ...O_t, q_t = S_i \mid \lambda)$ oznacza prawdopodobieństwo wygenerowania częściowej sekwencji obserwacji $O_1O_2 ...O_t$ do chwili *t* i osiągnięcia stanu S_i w chwili *t* przy danym modelu λ. $P(O \mid \lambda)$ jest wyznaczane w następujących krokach:

1. Inicjalizacja: dla $1 \le i \le N$ oblicz $\alpha_1(i) = \pi_i b_i(O_1)$.
2. Indukcja: dla $1 \le t \le T-1$ i $1 \le j \le N$ oblicz $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1})$.
3. Zakończenie: oblicz $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$.

Poprzez wyznaczenie wartości prawdopodobieństwa $P(O \mid \lambda)$ można określić, jak dobrze dany model pasuje do danej sekwencji obserwacji. W przypadku, gdy mamy dokonać wyboru najlepszego spośród kilku współzawodniczących modeli, jako najlepiej dopasowany do danej sekwencji obserwacji traktujemy ten, dla którego wartość $P(O \mid \lambda)$ jest największa.

3. ALGORYTMY EWOLUCYJNE W UCZENIU UKRYTYCH MODELI MARKOWA

3.1. Działanie algorytmu ewolucyjnego

Algorytm ewolucyjny jest to przybliżony algorytm optymalizacyjny, w którym stosowane są mechanizmy selekcji, rekombinacji i mutacji inspirowane przez biologiczny proces ewolucji. Algorytmy ewolucyjne zostały wprowadzone przez Hollanda w latach sześćdziesiątych ubiegłego wieku (Holland, 1975; Goldberg, 1995) i od tej

pory stosowane są z dużym powodzeniem dla rozwiązywania różnorodnych zagadnień: szeregowania i planowania procesów produkcyjnych, eksploracji danych (np. przewidywania i diagnozowania w medycynie), doboru parametrów procesów chemicznych, obrazowania i przetwarzania sygnałów oraz wielu innych.

Działanie algorytmu ewolucyjnego można opisać następująco: algorytm ewolucyjny rozpoczyna proces przeszukiwania od utworzenia populacji potencjalnych rozwiązań nazywanych osobnikami, które są reprezentowane przez chromosomy zawierające genetyczną informację o osobnikach. W każdym ewolucyjnym kroku, nazywanym generacją, chromosomy są dekodowane i ocenianie zgodnie z pewnym z góry przyjętym kryterium jakości nazywanym przystosowaniem, a następnie przeprowadzana jest selekcja w celu eliminacji osobników ocenionych jako najgorsze. Osobniki wykazujące wysokie przystosowanie podlegają mutacji (zmianie pojedynczych elementów w chromosomie) oraz przeprowadzanej przy pomocy operatora krzyżowania rekombinacji (wymianie pewnych fragmentów między chromosomami). Sama selekcja nie wprowadza żadnego nowego osobnika do populacji, tj. nie znajduje nowych punktów w przestrzeni poszukiwań, natomiast takie punkty wprowadzane są przez krzyżowanie i mutację. Dzięki krzyżowaniu ewolucyjny proces może się przesuwać w kierunku obiecujących obszarów w przestrzeni poszukiwań. Mutacja zapobiega zbieżności do lokalnego optimum. W wyniku działania operatora krzyżowania i mutacji tworzone są nowe rozwiązania, z których następnie budowana jest populacja następnej generacji. Warunkiem zakończenia algorytmu może być na przykład wykonanie pewnej określonej liczby generacji albo osiągnięcie zadawalającego poziomu przystosowania. Szczegółowy opis działania algorytmów ewolucyjnych można znaleźć w pracy (Figielska, 2006).

Ogólny schemat działania algorytmu ewolucyjnego jest następujący:

1. Ustaw $k = 0$.
2. Wygeneruj i oceń początkową populację $P(k)$.
3. Dopóki warunek stopu nie jest spełniony wykonuj:
   - 3.1. Ustaw $k = k + 1$.
   - 3.2. Wybierz $P(k)$ z $P(k - 1)$.
   - 3.3. Zmień $P(k)$ stosując operator krzyżowania i mutacji.
   - 3.4. Oceń $P(k)$.
4. Wyświetl najlepsze znalezione rozwiązanie.

3.2. Szacowanie parametrów ukrytego modelu Markowa

Jak już wspomniano wcześniej, jakość ukrytych modeli Markowa tworzonych przy użyciu metod klasycznych zależy silnie od jakości oszacowania początkowych parametrów modelu. Metody te wymagają również wstępnego określenia liczby stanów. (Np. przy rozpoznawaniu mowy (pojedynczych słów) przyjmuje się, że liczba stanów odpowiada liczbie głosek w słowie (Levinson, 1983) lub średniej liczbie obserwacji (każdy stan odpowiada przedziałowi obserwacji o długości ok. 10-15 ms.) w mówionej wersji słowa (Bakis, 1976); po wyznaczeniu wartości A, B i π, następuje korekcja modelu w celu określenia najlepszej liczby stanów (Kwong, 2001)). Zastosowanie algorytmów ewolucyjnych w trenowaniu ukrytych modeli Markowa umożliwia jednoczesną optymalizację wszystkich – łącznie z liczbą stanów – parametrów modelu, jak również stwarza możliwość utworzenia dobrej jakości modelu niezależnie od jakości wstępnego oszacowania jego parametrów.

W pracach (Kwong 2001; Figielska i Kasprzak, 2008) do uczenia ukrytych modeli Markowa stosowanych przy rozpoznawaniu pojedynczych słów, zaproponowane zostały algorytmy ewolucyjne, które w pojedynczym kroku optymalizują zarówno parametry A i B jak i liczbę stanów. Dla każdego słowa ze zbioru słów, które mają być rozpoznawane, tworzony jest osobny ukryty model Markowa (zakłada się, że jest to model typu „od lewej do prawej").

W zastosowaniach algorytmów ewolucyjnych do rozpoznawania słów chromosom jest kodowany jako ciąg liczb (genów), które odpowiadają wszystkim parametrom modelu (liczbie stanów N, prawdopodobieństwom przejść, a_{ij}, współczynnikom mieszania, c_{jm}, wartościom średnim, μ_{jm}, kowariancji, U_{jm}, a także liczbie składników losowej funkcji, M (przy założeniu ciągłej gęstości obserwacji, danej np. rozkładem Gaussa)). Parametry te zmieniane są w sposób bezpośredni w procesie optymalizacji.

Ocena populacji następuje przez wyznaczanie dla każdego jej osobnika g przystosowania f_g, które jest zdefiniowane następująco:

$$f_g = \frac{1}{K}\sum_{i=1}^{K} \log(P(O_i \mid \lambda_g)),$$

gdzie λ_g oznacza ukryty model Markowa reprezentowany przez chromosom g, K jest liczbą sekwencji obserwacji w zbiorze uczącym. Prawdopodobieństwo $P(O \mid \lambda)$ jest obliczane przez procedurę opisaną w rozdziale 2.3.

Wybór chromosomów rodzicielskich dla nowej populacji odbywa się w procesie selekcji, podczas której chromosomy zostają powielone w ten sposób, że osobniki o wyższym przystosowaniu mają większe prawdopodobieństwo wprowadzenia potomków do następnego pokolenia. W pracy (Kwong, 2001) została zastosowana

selekcja, zwana ruletkową, oparta na zasadzie proporcjonalności do przystosowania. Zapewnia ona, że liczba określająca ile razy dany osobnik jest wybrany, jest w przybliżeniu proporcjonalna do jego względnego przystosowania. Selekcję tę realizuje się przez symulację odpowiednio wykalibrowanej tarczy obrotowej (ruletki), gdzie każdemu chromosomowi z populacji odpowiada sektor o rozmiarze proporcjonalnym do wartości względnego przystosowania. Inny mechanizm selekcji, a mianowicie binarna selekcja turniejowa, został zastosowany w (Figielska i Kasprzak, 2008). Polega ona na wyborze spośród dwóch losowo wybranych osobników, jako rodzica, osobnika o wyższej wartości przystosowania. Zastosowanie selekcji turniejowej pozwala często uniknąć przedwczesnej zbieżności algorytmu, która może wystąpić przy stosowaniu selekcji ruletkowej, na skutek zbyt wczesnej (na początku procesu poszukiwań) eliminacji osobników zawierających „dobre podciągi", ale o stosunkowo niskim przystosowaniu.

Kluczowym elementem procesu optymalizacji są operacje krzyżowania i mutacji, które pozwalają na wprowadzenie nowej informacji do tworzonej populacji. Operatory te stosowane są z pewnym zadanym prawdopodobieństwem dla wyselekcjonowanych osobników rodzicielskich. Nowo utworzone osobniki potomne tworzą kolejną generację. Zastosowane w (Kwong, 2001) krzyżowanie polega na wymianie między osobnikami trzech losowo wybranych stanów. Krzyżowanie jednak, przy tworzeniu ukrytych modeli Markowa, nie wpływa zasadniczo na postęp procesu przeszukiwania. Główną rolę w tworzeniu nowych osobników odgrywa tutaj operacja mutacji. Jest ona rozdzielona na mutację zmieniającą liczbę stanów oraz mutację zmieniającą wartości pozostałych parametrów modelu: prawdopodobieństw przejść, a_{ij}, współczynników mieszania, c_{jm}, wartości średnich, μ_{jm}, oraz kowariancji, U_{jm}. Drugi ze wspomnianych operatorów mutacji dokonuje zmiany pojedynczego parametru x w chromosomie z pewnym ustalonym prawdopodobieństwem zgodnie ze wzorem $x' = x \cdot G(1.0, V)$, gdzie x' jest nową wartością parametru x, G (1.0,V) jest liczbą losową generowaną zgodnie z rozkładem Gaussa o średniej 1.0 i wariancji V.

Algorytmy ewolucyjne okazały się bardzo obiecującym narzędziem uczenia ukrytych modeli Markowa. W eksperymentach obliczeniowych badających efektywność modeli uczonych za pomocą algorytmów ewolucyjnych w rozpoznawaniu pojedynczych słów, uzyskano wysokie wartości stopnia rozpoznawania wynoszące ponad 95%.

W pracy (Kwong, 2002) przedstawiony jest algorytm ewolucyjny o charakterze dyskryminacyjnym, który w procesie uczenia bierze pod uwagę wszystkie współzawodniczące modele i minimalizuje błąd klasyfikacji dla całego zbioru danych uczących. Takie podejście poprawia możliwość rozpoznawania w stosunku

do sytuacji, gdy model jest tworzony dla każdego słowa z osobna. Algorytm ewolucyjny jest tu szczególnie użyteczny ze względu na stosowane kryterium oceny jakości (przy ograniczonym zbiorze danych mające postać kawałkami stałej funkcji parametrów klasyfikatora), które powoduje trudności w stosowaniu metod gradientowych. Stworzony dla 36 słów klasyfikator w eksperymentach obliczeniowych wykazywał błąd klasyfikacji o wartości około 9%.

Algorytmy ewolucyjne stosowane były również jako metody uczenia ukrytych modeli Markowa dla zagadnień rozważanych w bioinformatyce: do przewidywania drugorzędowych struktur protein (Thomsen, 2002; Won i in. 2005, Won i in., 2007), do analizy sekwencji DNA (Yada i in., 1994) oraz (Yada, 1995) w przypadku, gdy struktura ukrytego modelu Markowa jest reprezentowana przez drzewa stochastyczne.

Algorytm ewolucyjny został również wykorzystany do optymalizacji wartości początkowych parametrów ukrytych modeli Markowa (poddawanych potem reestymacji za pomocą algorytmu Bauma-Welcha) stosowanych do przewidywania cen na giełdzie papierów wartościowych (Hassan i in., 2007) na podstawie danych historycznych.

Głównym problemem, jaki pojawił się przy stosowaniu algorytmów ewolucyjnych w uczeniu ukrytych modeli Markowa, okazał się dość długi czas trwania procesu przeszukiwań konieczny do osiągnięcia zadawalających rezultatów. Wynika on z faktu przetwarzania przez algorytm dużych struktur danych – chromosom reprezentuje cały model lub nawet pewną liczbę modeli. W pracy (Kwong, 2001) w celu przyspieszenia procesu optymalizacji modelu, co kilka generacji uruchamiany jest klasyczny algorytm Bauma-Welcha (kilka jego iteracji) dla każdego chromosomu w populacji tak, że przystosowanie każdego chromosomu zostaje szybko poprawione. Działanie algorytmu ewolucyjnego można przyspieszyć także przez odpowiednie zaprojektowanie operatorów genetycznych, np. w (Figielska i Kasprzak, 2008) zrezygnowano z operatora krzyżowania, natomiast prawdopodobieństwo mutacji, duże na początku procesu optymalizacji, było stopniowo redukowane w miarę jego postępu. Niektórzy autorzy (Won i in. 2005) zaproponowali również ograniczenie nakładu obliczeń poprzez odpowiednie dobieranie struktury ukrytych modeli Markowa.

3.3. Przykład

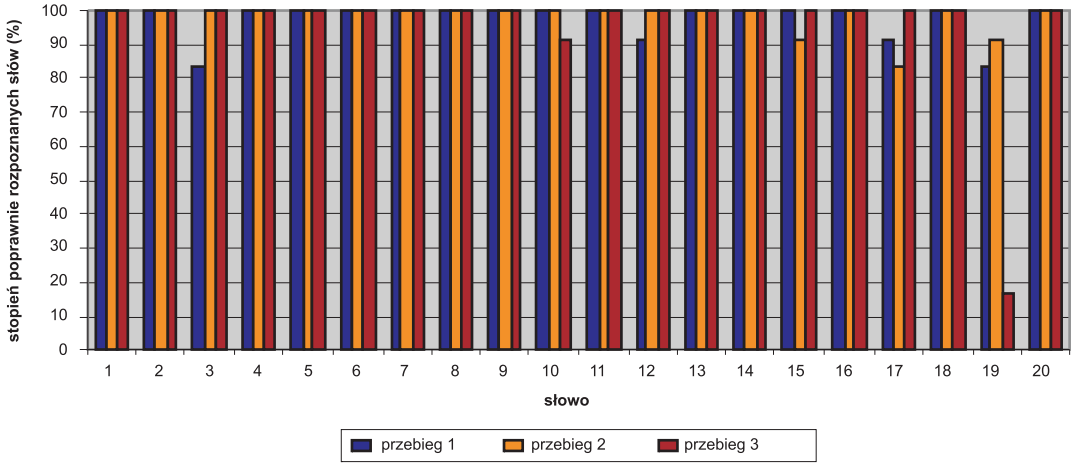
Poniżej przedstawione są wyniki eksperymentu obliczeniowego, w którym oszacowano stopień rozpoznawania słów osiągnięty przez ukryte modele Markowa trenowane za pomocą algorytmu ewolucyjnego. W eksperymencie wykorzystano 20 słów przedstawionych w Tabeli 1. Zbiór trenujący zawierał 18 wypowiedzi każdego słowa, podczas gdy zbiór testowy zawierał 12 innych wypowiedzi.

Tabela 1. Rozpoznawane słowa

numer	1	2	3	4	5	6	7	8	9	10
słowo	zero	jeden	dwa	trzy	cztery	pięć	sześć	siedem	osiem	dziewięć
numer	11	12	13	14	15	16	17	18	19	20
słowo	start	stop	lewo	prawo	góra	dół	puść	złap	oś	chwytak

Na rysunku 3 przedstawiony jest stopień poprawnie rozpoznanych słów uzyskany za pomocą ukrytych modeli Markowa uczonych w trzech przebiegach algorytmu ewolucyjnego. Dla poszczególnych przebiegów (uruchomień) stopień poprawnie rozpoznanych słów wynosił średnio:

- dla przebiegu 1: 97,50%
- dla przebiegu 2: 98,33%
- dla przebiegu 3: 95,42%



Rys. 3. Stopień poprawnie rozpoznanych słów uzyskany dla 3 przebiegów algorytmu uczącego

Na rysunku 3 można zauważyć, że większość słów była identyfikowana poprawnie. Niektóre słowa były jednak rozpoznawane gorzej niż pozostałe, np. słowo „oś" było identyfikowane jako „osiem", słowo „puść" jako „sześć", „dziewięć" jako „pięć".

4. PODSUMOWANIE

Chociaż do chwili obecnej główną rolę w trenowaniu ukrytych modeli Markowa odgrywają metody klasyczne, to coraz częściej pojawiają się prace proponujące stosowanie metod alternatywnych opartych na mechanizmach stochastycznych w tym algorytmów ewolucyjnych. Metody ewolucyjne są zdolne do tworzenia modeli

o optymalnych wartościach parametrów oraz dają możliwość stosowania kryteriów oceny modelu o dowolnych postaciach. Głównym wyzwaniem w przypadku stosowania algorytmów ewolucyjnych w uczeniu ukrytych modeli Markowa jest zmniejszenie wymaganego przez nie nakładu obliczeń, który jest konieczny do osiągnięcia dobrej jakości modeli.

Literatura

Baker, J.K., 1975. *The Dragon system – an overview*. IEEE Trans. Acoust., Speech Signal Process 23 (11), 23–29.

Bakis, R, 1976. *Continuous speech word recognition via centisecond acoustic states*, Proceedings ASA Meeting, Washington, DC.

Baum L.E., J.A. Egon, 1967. *An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology*. Bull. Amer. Meteorol. Soc. 73, 360-363.

Baum, L.E., T. Petrie, 1966. *Statistical inference for probabilistic functions of finite state Markov chains*. Ann. Math. Statist. 37, 1554–1563.

Baum, L.E., T. Petrie, G. Soules, N. Weiss, 1970. *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Ann. Math. Statistics, 41 (1), 164-171.

Baum, L.E., 1972. *An inequality and associated maximization technique occurring in statistical estimation for probabilistic functions of Markov process*. *Inequalities* III, 1-8.

Bladi, P., S. Brunak, 1998. Bioinformatics, the Machine Learning Approach. MIT Press.

Cai, J., Z.-O. Liu, 2001. *Hidden Markov models with spectral features for 2D shape recognition*. IEEE Trans. PAMI 23 (12), 703–713.

Chau, C.W., S. Kwong, C.K. Diu, W.R. Fahrner, 1997. *Optimisation of HMM by a genetic algorithm,* Proceedings ICASSP, 1727-1730.

Cheung, L.W.K., 2004. *Use of runs statistics for pattern recognition in genomic DNA sequences*, Journal of Computational Biology 11 (1), 107-124.

Coast, D.A., R.M. Stern, G.G. Cano, S.A. Briller, 1990. *An approach to cardiac arrhythmia analysis using hidden Markov models*, IEEE Transactions on Biomedical Egineering 37 (9), 826-836.

El-Yacoubi, A., M. Gilloux, R. Sabourin, C.Y. Suen, 1999. *An HMM based approach for off-line unconstrained handwritten word modeling and recognition*. IEEE Trans. PAMI 21 (8), 752–760.

Goldberg, D.E., 1995. *Algorytmy genetyczne i ich zastosowania,* WNT.

Figielska, E., 2006. *Algorytmy ewolucyjne i ich zastosowania*, Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki, 81-92.

Figielska, E., W. Kasprzak, 2008. *An evolutionary programming based algorithm for HMM training, in: Computational Intelligence: Methods and Applications*, L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, J. Zurada (eds), in series: Challenging problems of Science – Computer Science, L. Bolc (series editor), Academic Publishing House EXIT, 166-175

Hassan, Md. R., B. Nath, M. Kirley, 2007. *A fusion model of HMM, ANN and GA for stock market forecasting*. Expert Systems and Applications 33, 171-180.

Holland, J. H., 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.

Jelinek, F., 1976. *Continuous speech recognition by statistical methods.* Proc. IEEE 64, 532-556.

Kim, I., S. Chien, 2001. *Analysis of 3D hand trajectory gestures using stroke-based composite hidden Markov models*. Appl. Intell. 15, 131–143.

Karplus, K., K. Sjolander, C. Barrett, M. Cline, D. Haussler, R. Hughey, L. Holm, C. Sander, 1997. *Predicting protein structure using hidden Markov models*. Proteins: Struct., Funct. Genet. 1 (1), 134–139.

Kwong, S., C.W. Chau, K.F. Man, K.S. Tang. 2001. *Optimisation of HMM topology and its model parameters by genetic algorithms*. Pattern Recognition 34, 509-522.

Kwong, S., Q.H. Heb, K.W. Kua, T.M. Chana, K.F. Mana, K.S. Tanga Kwong S., C.W. Chau, K.F. Man, K.S. Tang. 2002, *A genetic classi_cation error methodfor speech recognition*. Signal Processing 82, 737 – 748.

Levinson, S.E., L.R. Rabiner, M.M. Sondhi, 1983. *An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition*, The Bell System Tech. J., 1035-1074.

Rabiner, L., 1989. *A Tutorial on Hidden Markov Models and selected Applications in Speec Recognition*, Proceedings of the IEEE 77 (2), 257-286.

Rabiner, L., Juang B., 1993. *Fundamentals of Speech Recognition*, Prentice Hall.

Thomsen, R., 2002. *Evolving the topology of hidden Markov models using evolutionary algorithms*. LNCS 2439, 861-870.

Won, K.-J.; T. Hamelryck, A. Prugel-Bennett, A. Krogh, 2005. *Evolving hidden Markov models for protein secondary structure prediction*. IEEE Congress on evolutionary computation 1, 33-40

Won, K.-J.; T. Hamelryck, A. Prugel-Bennett, A. Krogh, 2007. *An evolutionary method for learning HMM structure: prediction of protein secondary structure*. BMC Bioinformatics 8, 357.

Yada, T., M. Ishikawa, H. Tanaka, K. Asai, 1994. *DNA sequence analysis using hidden Markov model and genetic algorithm*. Genome Informatics 5, 178-179.

Yada, T., 1995. *Generation of hidden Markov model describing complex motif in DNA sequences*. IPSJ Trans. 40, 750-767.